

Ethics, Technology and the Challenges of Documenting History in Real Time

Howard Besser

Professor and Director, Moving Image Archiving & Preservation Program at New York University, and member of Activist Archivists, New York, USA
howard@nyu.edu

Sharon E. Farb

Associate University Librarian for Collections and Scholarly Communications at the UCLA Library, Los Angeles, California, USA
farb@library.ucla.edu

Todd Grappone

Associate University Librarian for Digital Initiatives and IT at the UCLA Library, Los Angeles, California, USA
grappone@library.ucla.edu

Ali Jamshidi

Digital Collections Specialist and Curator of the Green Movement collection at the UCLA Library, Los Angeles, California, USA
jamshidi@library.ucla.edu



Copyright © 2014 by **Howard Besser, Sharon E. Farb, Todd Grappone** and **Ali Jamshidi**. This work is made available under the terms of the Creative Commons Attribution 3.0 Unported License:

<http://creativecommons.org/licenses/by/3.0/>

Abstract

New technologies including mobile phones and use of social media software have made available a plethora of new sources for news and information that both complement and contrast with traditional news sources. This content is relevant to research libraries and archives around the world. Yet most of it does not get deposited into library and archival collections in traditional ways. Libraries and archives need to be innovative and proactive about seeking it out from numerous participants and scraping it off social network sites in real time to insure authenticity and reliability. At UCLA and NYU we have been collecting both digital and physical materials from the front lines of conflict and war-ravaged areas of the world. We've partnered with political activists to develop unique assemblages of ephemera collected on the front lines of social media revolutions. This paper will present two case studies that illustrate

the challenges and opportunities for collaboration and community engagement and utilizing new technologies including social media to capture and preserve history in real time.

Keywords: social media, citizen journalists, ethics, security and safety, role research libraries, archives, capture, curate and preserve historical record

Introduction

New technologies including mobile phones and use of social media software have made available a plethora of new sources for news and information that complement and contrast with traditional news sources. This content is relevant to research libraries and archives around the world. Yet most of it does not get deposited into library and archival collections in traditional ways. Libraries and archives need to be innovative and proactive about seeking it out from numerous participants and scraping it off social network sites in real time to insure authenticity and reliability. At UCLA and NYU we have been helping capture, curate and preserve both digital and physical materials from the front lines of conflict and war-ravaged areas of the world. We've partnered with political activists to develop unique assemblages of documentation collected on the front lines of conflict from around the world. This paper will present two case studies that illustrate the challenges and opportunities for international collaborations and utilizing new technologies including social media to capture and preserve history in real time.

The first section provides an overview and introduction to the issues. The second section presents two case studies that illustrate the challenges and opportunities for real time capture and preservation of social media of newsworthy events and activities. Section three discusses the technologies, methods and tools used to organize, capture and curate this disaggregated content and record of history. Section four discusses the legal, personal security and ethical challenges posed. Section five concludes with the role of libraries and archives to document and preserve the historical and cultural record in real time for present and future generations.

Documenting History in Real Time

Cultural heritage repositories have always collected material documenting history from a variety of sources. Among them are published newspapers from around the world, government documents, unpublished accounts from organizations and individuals, scrapbooks, letters, memorabilia, photographs, posters, handbills, leaflets and ephemera of all types. Reports are common about how technological innovation and broad public access to the Internet and social media have transformed opportunities and created new challenges to the

capture, preservation and long-term access of news content and the historical record.¹ When the historical record was primarily print based, long-term access required preservation of print media that was periodic, fixed, linear and typically singular rather than having distributed presentation and output. The historical and cultural record today and in the future relies increasingly on born-digital media requiring a fundamental change in methods used to capture, curate and preserve today's and tomorrow's cultural and historical record. The 24-hour news cycle now operates on a continuous loop of complex and distributed processes to capture, process, version, present, revise, respond and update.

Previously, most physical content and collections entered the repository as part of a large collection that follows some kind of internal organizational and retrieval scheme (such as a newspaper's photo collection with its own specific numbering and metadata scheme or an individual's file folders with labeling reflecting that individual's own personal logic). Libraries generally put this type of material in special collections and archives departments to help indicate that each collection reflects the organizational schema of the donating individual or organization. And archives create a unique finding aid for each collection because each collection's organizational scheme is so different that the archive could not create an organizational scheme that worked across multiple collections.

In the late 20th century as we moved more towards a world of born-digital collections, many challenges for cultural repositories (such as how to collect email correspondence or how to organize the folders on someone's hard disk) emerged. But for the most part, a repository could expect that all the material coming from an organization or individual would follow a single organizational logic and would be encoded in a particular set of file formats (such as all documents in a particular version of Microsoft Word for each time period and all digital movies in a particular version of Quicktime for each time period).

Like the Arab Spring and the Green Movement that preceded it (and inspired it), the growth of the Occupy movement was fueled by communication mediated on the Internet. But, partially because of the high level of broadband Internet access and the ubiquity of smart phones in the US, the number of digital photographs and video and audio recordings that movement participants posted online was astounding. Statistics from the photographic posting service Flickr show that six months after the first Occupy demonstration, more than half a million

¹ Abston, Jason. "Impact of Technology Upon the Newspaper Industry." 17 Feb. 2013. Web. <<http://bluecollarchronicle.com/impact-of-technology-upon-the-newspaper-industry/>>Abston, Jason, February 17, 2013. "Impact of Technology Upon the Newspaper Industry". *Blue Collar Chronicle*. Available Online bluecollarchronicle.com/impact-of-technology-upon-the-newspaper-industry/. Edmonds, Rick, Emily Guskin, Amy Mitchell, and Mark Jurkowitz. "Newspapers: Stabilizing, but Still Threatened." *The State of the News Media in 2013: An Annual Report on American Journalism*. Washington, D.C.: Pew Research Center Project for Excellence in Journalism, 2013. Edmonds, Rick, Guskin, Emily, Mitchell, Amy, Jurkowitz. July 18, 2013. "Newspapers: Stabilizing, but Still Threatened". Pew Research Center Project for Excellence in Journalism. *The State of the News Media in 2013: An Annual Report on American Journalism*. Center for Research Libraries. April 27, 2011. . *Preserving News In the Digital Environment: Mapping the Newspaper Industry in Transition*. April 27, 2011.

individual photos had been posted to this service with the tag of “#Occupy”.² Tens of thousands of individual videos were posted to YouTube during the first few months of the movement. By six months after the first Occupy action, 169,000 individual postings to YouTube had been tagged with “#Occupy”.³ The sheer volume of content created and the dissemination through commercial websites posed fascinating and time- sensitive problems for libraries and archives interested in preserving this material.

Many of the past artifacts collected by libraries and archives (such as letters and email and other types of correspondence, posters, handbills, etc.) have been replaced in today’s world by more atomized digital artifacts (such as blogs and social network postings) that are more connected to works originated by others. And in a world characterized by networked information and collaborative authorship, the cultural institution collecting material of historical and social importance will need to collect material coming from a variety of different sources, each with its own conception of metadata and file format standards.

In order to preserve this type of material, libraries and archives need to find smart ways to harvest metadata and analyze files as well as to influence the behavior of potential contributors. A number of the methods that might be useful for future user-contributed collections were explored in the projects of the Activist Archivists, which are outlined in this paper. Many of these methods are based upon the findings in prior projects on preserving born-digital material that Activist Archivist individual members had participated in. From the InterPARES II project (http://www.interpares.org/ip2/ip2_index.cfm) (2002-2006), we learned that if we hope to preserve electronic records, archivists need to be involved early in the life-cycle of that record, long before the record enters the archive. From the Preserving Digital Public Television project (<http://www.thirteen.org/ptvdigitalarchive/>), we learned the effectiveness of automating metadata collection from the moment of first recording.⁴ This paper builds on the work referenced and cited throughout and is itself a work in progress.

² March 24, 2012 Flickr statistics show 632,089 items tagged with “#Occupy”, 164,304 tagged with “Occupy Wall Street”, 179,454 tagged with “Occupy Protest”, 113,904 tagged with “#OWS”, 40,572 tagged with “Occupy Movement”, 27,202 tagged with “Occupy Oakland”, and 9,164 tagged with “Zucotti Park” (location of the first NYC occupation).

³ March 24, 2012 YouTube statistics show 169,000 items tagged with “#Occupy”, 98,400 tagged with “Occupy Wall Street”, 70,500 tagged with “Occupy Protest”, 50,300 tagged with “#OWS”, 54,800 tagged with “Occupy Movement”, 13,400 tagged with “Occupy Oakland”, and 6,690 tagged with “Zucotti Park” (location of the first NYC occupation).

⁴ Besser, Howard and Kara van Malssen. Pushing Metadata Capture Upstream into the Content Production Process: Preliminary Studies of Public Television, DigCCurr 2007: an international symposium in digital curation, Chapel Hill, NC, USA. 18-20 April 2007. Available online <http://ils.unc.edu/digccurr2007/papers/besserVanmalssen_paper_4-1.pdf>.

Partnering With Activists I: The Green Movement—Iranian Election 2009

One of the methods to capture and curate news and historical events in real time is to collaborate with those on the ground closest to where the events are occurring. In some cases, without an insider -- or in today's vernacular, citizen journalism or public journalism -- it would simply not be possible to obtain the on the ground information and perspective and along with documentation otherwise. For academic and research libraries, collaborating actively with those making as well as recording historical events is a relatively new practice. Historically, academic libraries have relied on international acquisitions of the published record (previously print newspapers, journals and other material) collected and curated by subject specialists and faculty who are typically studying history and current events and conflict rather than making of history. The two cases studies included in this section highlight the issues faced by grass roots activists in collaborating to capture and preserve history in real time.

The Green Movement Collection at UCLA is a part of a larger UCLA Library international collecting and collaboration initiative. UCLA like NYU is a global research university with increasing flow of students, faculty and researchers around the world. In addition, in both cases our local communities, Los Angeles and New York, reflect and are composed of the larger global communities, with each having the largest population of Iranians outside of Iran and Armenians outside of Armenia, Central Americans outside of Central America, etc. The imperative for community as well as international engagement and collection building is integral to the mission of higher education and research libraries and archives.

The Arcadia Fund generously funds the International Digital Ephemera Initiative (IDEPIDEI) at UCLA. The IDEPIDEI is designed to collaborate with international partners to capture at-risk digital ephemera from around the world that will be freely and openly available to all. UCLA's first collaborator is the National Library of Israel. The IDEPIDEI currently consists of the Tahrir Square collection, the Ghezi Park collection and the Green Movement. The Green Movement collection, which is still growing, currently consists of the following born-digital content: more than 5,000 videos in the categories of protests, interviews and clips, Rasa TV; more than 8,750 images; more than 730 posters; two websites; approximately 1,200 issues of underground newspapers; and miscellaneous files.

The UCLA Library was fortunate to meet Ali Jamshidi, who is currently Digital Collections Specialist and curator of the Green Movement collection at UCLA. Ali was previously the founder and leader of one of the most important social media platforms for distribution of information about the Green Movement and post-election protests and founder of the prominent Iranian journalistic outlet Tahavole Sabz. Below are excerpts from Ali's first-person account of his role as activist and current archivist at UCLA that illustrate a couple of the ethical and security risks inherent in capturing history in war-torn areas of the world in real time. :

In June 2009 the Iranian people came to the streets demanding to have their voice heard. It was a month I can never forget, and a month that changed my life forever.

When I saw my people beating their chest demanding respect and dignity, I could not stay silent. I jumped to my computer and established one of the largest underground networks of student and political activists inside Iran. My network extended through all of Iran's major cities: Tehran, Shiraz, Mashhad, Tabriz and Isfahan.

As the political situation got worse and people were beaten and shot in the streets, my network grew bigger and bigger and I became one of the key leaders in arranging protests inside of the country. My statements and articles were published, downloaded and printed tens of thousands of times inside of the country for distribution to students and citizens.

During these events I was outside of Iran. At first, I was in hiding in Turkey. From there, I fled illegally to Cyprus waiting to get a visa to a safe and free country. The story of my journey to Cyprus is a long and interesting one, and I hope in the future you can read about it in my book.

Serious security concerns forced me to flee Turkey for Malaysia and apply to the UN as a refugee. My activities did not stop during my stay in Malaysia, as my network expanded and grew to a level which was unmanageable by Facebook alone. I founded the website Tahavole Sabz, which means "Green Evolution". The site became one of the major Green Movement websites and one of the official news sources from 2010-2012. In addition, the Green Movement under Mr. Mousavi and Mr. Karroubi launched an official satellite TV channel – RASA TV – which I was one of two hosts for every night. At RASA TV I also launched my own special program called "Face to Face" every Thursday evening.

The high level of my work did not go unnoticed. The regime was enraged and published dozens of angry articles across their propagandist machine about my network and activities.

During my stay in Malaysia, the regime website Raja News, close to the Revolutionary Guard, published a highly publicized article accusing me of working for the U.S. government, MI6 and the Mossad. My association with these major intelligence agencies was news to me!

My refugee application was eventually approved by the U.S. government and I then spent my days involved with my network and activities while hoping and praying for a flight date. This was a difficult time for my mental health, as every morning I would wake up hoping that I would get a simple call from the UN letting me know that I was free to go to a safe country.

It would be another three months, waiting for the call every day, until finally I was told in December 2011 that I would depart Malaysia for the United States in February or March 2012 another two or three months of hell, living in cramped quarters spending my days glued to the computer screen witnessing the brutality of a regime that no longer valued the lives of its own citizens.

Just when I thought that there was an end in sight, I received a telephone call on a morning in February 2012. It was a call from a man who said he worked for Bukit Aman, the Malaysian Special Intelligence Police, asking me to appear at their offices tomorrow morning at 9 AM.

I was terrified, but I had to go. I was interrogated five times by Bukit Aman officers and many times by telephone after that. They asked who supported me financially, who I was working with in Malaysia, the extent of my Malaysia network, which organizations and individuals in the US are helping me and more. I did my best to protect my friends and networks, but Bukit Aman had compiled a large intelligence file on me including pictures they took of me meeting others in public without my knowledge or consent.

Finally, on March 19, 2012, I flew to Los Angeles and entered the United States on Iranian New Year. It was such an amazing feeling to celebrate a new year in a free country with no stress after four years of pain.

In reality, I was about to enter a new stressful time. When I was in Malaysia I thought that all of my problems would be solved when I got to the United States. When I got here I realized that some problems were solved and new ones had emerged. I didn't know the language, the geography, the culture. It was like I came from the moon and was put here. I couldn't find work, even at grocery stores, because I lacked the language and financial and legal background here. It was difficult to relive my refugee experience and explain it to each employer when they asked why I had no history here.

Despite this, I was walking with a friend around UCLA and we saw an announcement for the Arab Spring ephemera project for Egypt. My friend encouraged me to submit a proposal about my personal collection and Green Movement experience to the UCLA Library. I created a great archive of more than 3,000 videos sent to me directly by my network inside Iran. I also had dozens of Facebook groups, website files, posters, photographs, underground articles, statements, underground newspapers and more.

In the time I was waiting to get an answer from UCLA, my money was running out, and I had no immediate employment opportunities. In despair I discovered free computer certification courses for refugees being offered at a community college in Houston.

I took the last of my money and got on a flight for Houston. I enrolled in the program and started to work at McDonalds but was placed in the front to serve customers. I could not understand what people were saying and was quickly forced to leave, ending up working at a beer cooler factory.

After few months, when I was back in LA, I was contacted by UCLA, and very quickly I started my job as a digital collection specialist. What I brought to UCLA was a history of my people fighting in the street for their voice to be heard, and I believed it should be captured and saved in a safe place for history and for future generations to see what their parents and grandparents fought for. It doesn't matter if the movement succeeded or not, but what matters is the millions of people who tried to do something,

to make a difference in their world and the world at large. They were arrested, beaten, tortured and killed. They tried to make a better Iran, and I was able to capture some of their struggle. I thank the UCLA library for its commitment to making sure their voices are heard and preserved.

The UCLA Library thanks Ali Jamshidi for his fearless dedication and ongoing work to capture, curate and preserve history in real time.

Partnering with Activists II—The Occupy Movement⁵

The Occupy movement that emerged in the US in the fall of 2011 grew rapidly by posting both text and audiovisual works on online sites that would likely disappear. Archivists primarily from New York University's Moving Image Archiving & Preservation program worked with Occupy activists to assure greater longevity for the photos, audio and video recordings of the movement and its events. Their work offers important lessons in documenting history in real time, particularly with the advent of ubiquitous cellphone technology and social networks.

The Occupy movement began in September 2011 in New York City and quickly spread to cities and towns across the US (and eventually to other parts of the world).⁶ The movement's key slogan – “We are the 99%” – reflects that the movement was fueled by a moral outrage at the control exerted on society by a small minority of the populace. The movement's name – “Occupy” – points to its tactic of “occupying” public physical spaces for 24 hours per day seven days per week both to highlight the importance of those spaces to society's discourses and to maintain a constant presence where people who pass by cannot help but notice the movement. This 24/7 presence in physical space also led to the development of self organizing and community building within the movement itself and is reflected in the communal feeding of large numbers of participants (numbering in the hundreds, or in the case of NYC sometimes numbering in the thousands) and in collective providing of services for all participants (in the form of lending libraries, electrical power, wireless internet services, etc.).

In addition to physically occupying key public spaces, the movement engaged in extensive large-scale demonstrations involving thousands or tens of thousands of participants. Often these demonstrations highlighted what the movement saw as particular examples of systemic problems in society – the government's bail-out of financial firms (while not rescuing the worst-off individuals), the seizure of peoples' houses via foreclosure, etc. A major characteristic of the movement was the broad creativity shown in signs carried in protest marches and in creative street-theater, where protestors would dress as bankers or governmental officials and act out satiric scenarios.

⁵ For further details about the Activist Archivists' work with the Occupy Movement, see their website (<http://activist-archivists.org/>); Besser, Howard. "Archiving Aggregates of Individually Created Digital Content: Lessons from Archiving the Occupy Movement", *Preservation, Digital Technology & Culture*. Volume 42, Issue 1, Pages 31-37, March 2013; and <http://besser.tsoa.nyu.edu/howard/Papers/besser-girona-occupy-paper.pdf>.

⁶ According to Wikipedia, by October 9, 2011 (3 weeks after its beginning in NYC), Occupy protests had taken place in over 95 cities across 82 countries, and over 600 communities in the US. (http://en.m.wikipedia.org/wiki/Occupy_movement accessed August 19, 2012)

The material generated by the Occupy movement looks very much like the type of material that will be entering the archives and library special collections of the future. It is a vast quantity of user-generated everyday material created by a multitude of different users.⁷ There is no easy way to control for quality, file format or metadata. Unlike most organizational collections that try to enforce standards for metadata and file formats, there are not even guidelines suggesting what schemes should be followed. And because the content comes from so many individuals, it lacks even the semi-consistency that a single individual would apply to the items that he or she creates. And what might logically constitute a future Occupy media collection is actually found today spread over a multitude of commercial social networks (such as Flickr, YouTube, and Facebook) that each add their own organizational idiosyncrasies and offer no guarantee that the material will remain posted for any length of time.

In response to the Occupy movement, in October 2011 students and recent graduates of NYU's Moving Image Archive and Preservation Program (MIAP)⁸ began efforts to explore the archiving and preservation of the media being generated by the Occupy movement. They felt that much of the spirit, decentralization, self-organization, playfulness and whimsy of this protest movement would be lost to history if the media that documented this did not survive. Joined by MIAP Director Howard Besser, the group took on the name Activist Archivists and began work on about a dozen different projects to archive the born-digital media content related to this movement (<http://www.activist-archivists.org/>), with most of the projects having potential impact on the archiving and preservation of all types of material that might be collected by cultural repositories in the future.

Many of the sub-projects involved collaboration with various partners. These included both collecting institutions (such as the NYU Library's Tamiment Collection) and "working groups" from the Occupy Wall Street movement (including both the "Archives" working group, which mainly dealt with collecting non-digital artifacts such as posters and signs, and the "Media" working group).

It is important to note that certain predispositions of the Occupy movement may not be relevant to libraries and archives building collections from other sources. Those in the Occupy movement were very suspicious of conventional organizations, including universities and libraries, and often needed convincing that a conventional cultural institution might be a good repository for the artifacts that they created. Occupiers could also be characterized as having a "do-it-yourself" mentality, not wanting to rely on professionals outside their community to organize and provide access to the material. This was part of a critique of conventional media dissemination outlets, which appeared to not do a good job of explaining the movement and appeared to manipulate news coverage. The Occupiers wanted to control their own story. Their ideology also made them suspicious of any type of exclusive arrangement, including giving their material only to a particular repository. And their consensus decision-making process

⁷ Besser, Howard. *Amateur Collections and Scholarship: Lessons from the impact of amateur collections on Cinema Studies research and on Film Archives' practice*. Photo Archives and the Photographic Memory of Art History, Part III, Institute of Fine Arts, New York University, NY, NY, USA. 25-26 March 2011.

⁸ <http://www.nyu.edu/tisch/preservation/>

made it difficult for a repository to try to come to an agreement with the group, as a group discussion on a topic such as this might range over several meetings, and each meeting might be composed of a slightly different group of participants (and discussion from previous meetings had to be repeated to and accepted by the newcomers).

Because many in the Occupy movement did not immediately recognize the value of saving artifacts representing their movement, one of the first projects undertaken by the Activist Archivists was to create a short, simple postcard that succinctly explained why saving these artifacts should be important to the movement. Much effort was made to explain this using the value system of the movement, and significant effort went in to avoiding what many in the movement referred to as “outsider language”. The “Why Archive” postcard briefly explained that saving this material would serve values dear to the movement, including accountability, self-determination, sharing, education and providing continuity. It explained that it was important to “Record and Collect” what was happening in the movement in order to “Preserve the record”.

Another important Activist Archivist project sought to provide advice to those recording Occupy events. The “7 Tips to Ensure our Video is Usable in the Long Term” offers important advice to those recording video, still images or sound. Advice includes collecting details about the recording, keeping the original raw unaltered footage, making the recording discoverable/accessible, contextualizing it, making it verifiable, allowing others to collect and archive the recording or being very careful about archiving it yourself. (For this last point, the tip suggests looking at the Library of Congress’ website on personal archiving.) Ideas from these “7 Tips” were expanded into a much lengthier set of advice in “Best Practices for Video Activists.” This document also discussed legal restrictions involving getting permission from those you record (which in the US differs depending on which state the recording takes place in) as well as issues involving copyright (and stresses the idea of executing a Creative Commons license, which will allow a repository to archive the material and make it available in the future).

An Activist Archivist research project revealed significant issues for any organization seeking to preserve digital videos posted to commercial websites. MIAP student Rufus de Rham’s “Metadata Loss During Upload or Download” empirical study uploaded digital video files from three different consumer recording devices (iPhone, Android and Canon t2i) to four different Internet dissemination services (YouTube, the Internet Archive and both Vimeo’s paid and free services). He then examined more than 200 fields of metadata present in the original files and found that YouTube and Vimeo’s free service stripped out almost all the metadata and only the Internet Archive’s service maintained all the metadata intact. Important metadata such as date, time and GPS location were stripped out of the file header by YouTube and Vimeo’s free service. This means that cultural organizations seeking to collect videos posted to online services would have to reconstruct date, time and location information for anything they collected from YouTube or Vimeo’s free service.

Activist Archivists also worked on a “Best Practices for Content Collectors.” This document aimed to pull together information gleaned from other Activist Archivist projects and to offer advice to cultural repositories on how best to collect this type of content. It suggested that with content pulled from the Internet Archive, metadata for date and time recorded (and possibly geographic location as well) could be automatically extracted from metadata in the file headers. It discussed some Occupiers’ sensitivity around traditional organizations and exclusivity and suggested ways in which an archive could collect materials without being tied to an exclusive agreement. It discussed the sensitivity of some protesters around security and offered suggestions for using software tools that could automatically detect faces and place small black bands over peoples’ eyes in order to make individual identification difficult.

Two Activist Archivist members worked closely with NYU’s Tamiment archival collection, which had arranged for the digital audio recording of daily two-hour “Think Tank” discussions of strategies and tactics by Occupy Wall Street participants. The Activist Archivists suggested ways in which important metadata could be automatically extracted from the recordings as well as working on the spreadsheets of metadata related to each recording. Guidelines for the recordings insisted that key metadata be captured in redundant ways. For example, the date that should have been captured by the Zoom H2n digital audio recording device was also embedded within the file-name and was supposed to be orally recited as part of an initial script that was read into the recorder at the start of each session. This redundancy was important because the recording device’s date and time were not necessarily reset when the device’s battery ran out of energy, and the oral script was not always read at the beginning of the recording. This form of redundancy would also protect against a future time when the metadata might become separated from the content. And the reading of the same exact script at the beginning of each session would allow for future, more sophisticated speech recognition software to automatically extract important metadata. The script also is designed to systematically record information including who is responsible for the recording, what is expected to be discussed in this session, and the hat participants’ desire to execute a Creative Commons license that will allow archives to preserve and disseminate the recording.

Activist Archivists also advised NYU’s Tamiment archival collection on preserving digital videos posted on YouTube. Selection of which videos to capture and preserve had been a tedious process, with the Tamiment director viewing each YouTube video and deciding whether or not it was worthy of preservation. This process would not scale to the over 100,000 videos that had been posted to YouTube tagged with “Occupy Wall Street” in the first just few months of into the movement. Activist Archivists suggested that categories for the videos be developed (celebrity visits; internal workings such as the library, kitchen and media; confrontations with police; labor union involvement; housing/foreclosure protests; etc.) and that selection of the most important videos be crowd-sourced to members of the Occupy Wall Street working groups (both archives and media). Each participant could fill out a web form indicating what they thought were the five most important YouTube videos to preserve in each category. This method would scale up as the number of videos posted to YouTube increased, and it was more in line with the participatory nature of the Occupy movement. This method was not implemented due to the prolonged illness and eventual death of the Tamiment director.

Activist Archivists tackled a wide variety of problems inherent in selecting, capturing and preserving media related to the Occupy movement. They explored various methods for convincing the individuals who recorded these events to make sure that their recording devices were set to automatically record date, time and location; to use consistent metadata and file-formats; to post to sites that will not throw away their metadata; and to execute Creative Commons licenses that will give cultural repositories the legal right to preserve and make available these recordings. They also advised collecting institutions on how to articulate their need to preserve to the individuals that make recordings and provided advice on both redundant methods of capturing metadata for recordings and on scaling the selection process in creative ways. These efforts are likely to prove useful for solving a problem that most collections will face in the near future: how to organize, preserve, and provide access to that large amount of user-generated content that most collections will receive in the future and not have the time to catalog or convert.

Technologies, Methods and Tools to Organize, Capture and Curate History in Real Time

The human connections gave us the opportunity to collect content relevant to the movements discussed. In some cases these collections came to the library already curated. Content such as websites, YouTube videos and social media of the movements were collected as primary sources and exemplars for the type of media they had collected. These collections were captured using standard software's such as the Web Archiving Service from the California Digital Library and with homegrown software, the Twitter Vacuum and YouTube capture-and-indexing software developed by the UCLA Library. The video tools developed by UCLA Library as part of our NewsScape project allow for fulltext search, visual browse from still captures and named entities among other things. These collections were captured and described using established best practices. In the case of RasaTV, the Green Movement's broadcast news-style YouTube channel, the video had to be translated and time coded to make the content keyword and entity searchable. Given the size and scope of the collection, the challenge to the Library was to accurately describe the content in both time and place. In the case of the Green Movement content, we have used the originating collector, Ali Jamshidi, for description and context setting for the videos. Due to the fact that a good deal of our content was given to Ali by eyewitnesses, we have used their original email detailing the time, place and activity. Where metadata existed, we have incorporated that into our cataloging procedures. We have augmented traditional cataloging with native speakers to translate and provide subject descriptions where possible. These descriptions from the originating capture agent have proven invaluable for this style of collecting. As outside observers to conflict, our mindset on the activity that has been captured is often subjected to outside influences from places like the news media, which is not always accurate. As both Ali's case study of the Green Movement and the Occupy case study reflect, traditional media do not always accurately show the events or context for the events as they are occurring. Setting context for the content in these types of collections is important, as non-activist agents will often use similar footage from the same protest to impart a different message. These types of collections are built to study content from the perspective of humanistic inquiry – that is, inquiry based on objectivity rather than interpretation and inference.

As activism has moved online, our opportunity to collect has greatly increased as we have described in this paper. We would also like to note that in addition to the first-person collections, and our activist-gathered content, we have also been collecting ephemeral content related to the supporting Internet activism. We have collected not only websites but software packages and server configuration files as well. These collections include items such as web capture before and after sites have been hacked and “distributed denial-of-service” attack logs.

All of this collecting does require us to be vigilant with regard to the subjects (people) in our collection. Contributors run the risk of being jailed or even killed. In some cases we have had to restrict access to these collections. We anonymized content where we can to allow for open access. For example, we have parsed and to some extent masked the Internet Protocol octets of the sites' visitors, leaving geographic relevance. While we make more content available, we strive to ensure privacy by using facial recognition software to identify and blur the actual identity of the faces. While not entirely accurate, we do both automated and human generated quality assurance to ensure that collections are vetted before they are presented on the open web. The most important factor in this type of collecting is trust. We make every effort to be seen as trustworthy and to reassure our subjects that we are not going to mishandle content that means so much to them personally.

Legal, Ethical and Personal Safety and Security Challenges

The Universal Declaration of Human Rights (UDHR) provides an international legal framework and context for questions regarding freedom of expression and freedom to receive and impart information. Specifically, Article 19 provides: “everyone has the right to freedom of opinion and expression; this right includes the freedom to hold opinions, without interference and to seek, receive and impart information and ideas through any media and regardless of frontiers”..⁹ Proclaimed in 1948 by representatives with diverse legal and cultural backgrounds from all over the world, the UDHR sets out “fundamental human rights to be universally protected.”¹⁰ In addition, the Geneva Conventions and Additional Protocol I, which has jurisdiction over international armed conflict, may provide some protections. However, as Emily Crawford points out, “the increasingly amorphous character of 21st-century armed conflicts has meant that journalists most often find themselves reporting on non-international armed conflicts, or conflicts that do not meet the threshold of armed conflict under international law.”¹¹

⁹ United Nations Human Rights, *Universal Declaration of Human Rights*. Web. 2014. <>.

¹⁰ <http://www.ohchr.org/en/udhr/pages/introduction.aspx>

¹¹ Crawford, Emily. August 30. 2012. “The International Protection of Journalists in Times of Armed Conflict and the Campaign for a Press Emblem.” *Sydney Law School Research Paper* No. 12/61. 30 August 2012. Web. <<http://dx.doi.org/10.21.39/ssm.2139153>>

State and country legal protections for freedom of speech, press and information are also varied and diverse and subject to ongoing political change. In the United States, there is a long-established legal tradition of freedom of speech, freedom of the press and freedom of information that is explicitly embodied in the Constitution and First Amendment of the Bill of Rights, which reads:

Congress shall make no law respecting an establishment of religion, or prohibiting the free exercise thereof; or abridging the freedom of speech, or the press; or the right of the people peaceably to assemble, and to petition the Government for a redress of grievances.¹²

Privacy protections and confidentiality of sources are another legal issue raised in the context of capturing, curating and preserving historical events in real time. There is no shortage of examples of where reporting or documenting current events in real time can have significant impact and personal risk to those involved. In 2007 the International Press Centre's *Killing the Messenger: report of the global inquiry by the International News Safety institute into the protection of journalists* found "over one thousand journalists and support staff have died trying to report the news around the world in the past 10 years; an average of two a week" and "at least 657 men and women were murdered—eliminated as they tried to shine light into the recesses of their societies—and only one in eight of the killers were prosecuted".¹³

Legal issues regarding international jurisdiction also come into play, as Lee Salter points out in his chapter on *Indymedia and the Law: Issues for Citizen Journalism*: "online citizen journalism presents us with a conceptual conundrum; a citizen is the subject of a state, but the internet allows material to transcend the jurisdictional boundaries of the state."¹⁴). As Salter points out, a citizen journalist could write copy in one state and jurisdiction, upload and publish it in another and upload and host it in still other servers in other countries. In this highly distributed network, how and what jurisdictions are accountable and responsible, and what processes exist or must be developed to address and appeal? Issues of jurisdiction may include issues of due process, so it is critical going forward that processes are developed by those involved that follow international guidelines such as the UDHR and UNESCO's Civil Society's Joint Statement for Post 2015¹⁵.

¹² U.S. Constitution. Amend. I.

¹³ International Press Centre. *Killing the Messenger: report of the global inquiry by the International News Safety institute into the protection of journalists*. 2007. Web. <<http://>>

¹⁴ Salter, Lee, "Indymedia and the Law: Issues for Citizen Journalism", in Stuart Allan, ed., Einar Thorsen, ed., *Citizen Journalism: global perspectives*. Peter Lang International Academic Publishers, 2009.

¹⁵ United Nations Educational, Scientific, and Cultural Organization (UNESCO). World Press Freedom Day 2014. *Civil Society's Joint Statement for Post 2015*. Available Online: www.unesco.org/new/en/unesco/events/prizes-and-celebrations/celebrations/international-days/world-press-freedom-day/background-materials/civil-societys-joint-statement-for-post-2015/

Since 1993 the United Nations General Assembly proclaimed May 3 as World Press Freedom Day to “serve as an occasion to inform citizens of violations of press freedom—a reminder that in dozens of countries around the world, publications are censored, fined, suspended and closed down, while journalists, editors and publishers are harassed, attacked, detained and even murdered.”¹⁶

In *The Book of Laughter and Forgetting*, Milan Kundera writes, “the struggle of man against power is the struggle of memory against forgetting.”¹⁷ Two recent legal cases in the news are pertinent to this discussion. First is the recent international case involving subpoenaed transcripts and records from Boston College in the U.S. by British and Irish governments. This case has huge implications for any institution interested in collecting contested historical events while witnesses are alive and concerned over protecting their personal safety and security as well as the integrity and accuracy of the historical record. Second is the recent Court of Justice of the European Union ruling that Google and other search engines must remove certain links if users complain, regardless of whether the links to articles are facts or true or not. As of this writing, it is unlikely that this ruling would affect search engines in the U.S. and elsewhere; however, its implications are international and wide reaching. This matter originated when a Spanish businessman complained that an article about him from a decade ago could potentially cast a negative light on his business reputation in the future. He did not claim the article or any facts were false or misleading but instead merely that it was “embarrassing”.¹⁸ If this ruling is allowed to stand or expand in any way, it will have the effect of censoring the historical and cultural record and undermine the mission of libraries and archives to collect and preserve.

Role of Libraries, Archives and Museums Documenting the Historical Record in Real Time

In *The Idea of the University: A Reexamination*, Jaroslav Pelikan describes the four-pronged mission of higher education: “1) the advancement of knowledge through research, 2) the extension of knowledge through teaching, 3) the diffusion of knowledge through publishing and 4) the preservation of knowledge through libraries, archives and museums”.¹⁹ Robert D. Leigh describes the primary mission of libraries and archives of all kinds, which “during centuries of their existence have had a common objective—one so generally accepted that it is seldom made explicit. It is the conservation and organization of the world’s resources of recorded thought and fact so as to make them available for present and future users” (Leigh, 1950).²⁰

¹⁶ "World Press Freedom Day 2014." . UNESCO, 3 May 2014. Web. . <<https://en.unesco.org/events/world-press-freedom-day-2014>>.

¹⁷ Kundera, Milan. *The Book of Laughter and Forgetting*. 1994. Translated from the French by Aaron Asher. New York: Harper Collins, 1996.

¹⁸ Gierhart, Cindy. “European Court Rules Google Must Remove Links at User’s Request.” *Reporters Committee For the Freedom of the Press*. 13 May 2014. Available Online: <http://www.rcfp.org/browse-media-law-resources/news/european-court-rules-google-must-remove-links-users-request>

¹⁹ Pelikan, Jaroslav. *The Idea of the University*. Princeton, New Jersey: Yale University Press, 1994.

²⁰ Leigh, Robert, D. *The Public Library in the United States*. New York: Columbia University Press, 1950.

Included in the mission and vision of the National Archive and Records Administration of the U.S. is the explicit acknowledgement of the vital role that records play in a democracy.²¹

The role of libraries and archives to capture and preserve the historical record is well established. As evidenced above and throughout, new technologies continue to present both opportunities and challenges to fulfilling that fundamental mission. In addition, new technologies present new ethical and legal challenges regarding access and preservation as well as privacy and confidentiality of user information and whether institutions can adequately protect human subjects, activists or citizen journalists from challenges presented by law enforcement, as the recent Boston College oral history case reflects. It is imperative that collecting and preserving institutions and organizations directly address the challenges to protecting the confidentiality of informants and to proactively inform all parties of the limitations or challenges so they can make an informed choice as to risk. Christine Anne George calls for an archival call to action “archival privilege,” which the authors of this paper would expand to include libraries’ and museums’ ability to invoke a “cultural heritage organizations privilege” to protect information and sources and to provide immunity against subpoenas issued to require disclosure of information regarding sources.²²

In closing the authors offer a few suggestions for research libraries engaged in utilizing technology to collect and preserve the historical record in real time. First, be proactive. Technology and current significant events change rapidly, and important ephemeral information disappears. It is no longer possible to merely take a passive or deliberate approach of relying on the published record alone. Second, wherever possible, a holistic approach is preferred. As with any event, there is value in multiple and diverse perspectives rather than a singular perspective and approach to documenting and presenting history and the historical and cultural record. Third, get as close to the source as possible. This has always been important, and especially wand where archival and manuscript collections and other ephemera reflect perspectives of actions, events and individuals that otherwise may not be heard. This is also where community engagement and collaborations with those closest to the events and activities present the greatest opportunities. Fourth, know the legal and policy context. As discussed above, there are serious legal and personal consequences of some of this work, and the local, regional and national context must be fully understood to protect those involved. Finally, research libraries engaged in this type of work should plan for training in legal, technical and cultural issues in order to build and curate content and collections that document history in real time. We posit that the collecting of 21st-century ephemera renders the ancient axiom that “history favors the victor” obsolete. This ephemeral content housed in open repositories

²¹ *The National Archives and Records Administration Fiscal Year 2014-2018 Strategic Plan*. Washington, DC: National Archives and Records Administration, 2014. Available Online: <http://www.archives.gov/about/plans-reports/strategic-plan/2014/nara-strategic-plan-2014-2018.pdf>.

²² George, Christine Anne. “Archives Beyond the Pale: Negotiating Legal and Ethical Entanglements after the Belfast Project”. *76 American Archivist* (2013). *SUNY Buffalo Legal Studies Research Paper* No 2013-40. 20 March 2013.

retains multiple perspectives on world events. And that is what will be the most compelling reason for collecting this material in the future.